融合全局与局部视角的光场超分辨率重建 *

邓 武，张旭东，熊 伟，汪义志

(合肥工业大学 计算机与信息学院，合肥 230601)

摘 要：针对光场相机结构和像素传感器分辨率的限制导致光场图像空间分辨率和角度分辨率都较低的问题，提出一种融合全局与局部视角的光场超分辨率重建算法，同时提高光场图像的空间分辨率和角度分辨率。首先根据待重建新视角的位置，自适应选择局部视角，利用空间超分辨率卷积神经网络提高全局视角和局部视角的空间分辨率，然后提取并融合全局视角和局部视角在新视角处映射图像的深度特征和颜色特征，通过角度分辨率卷积神经网络重建获得新视角图像。实验结果表明，与现有方法相比，峰值信噪比(PSNR)提高约 3 dB，结构相似性指数(SSIM)提高约 0.02，有效地解决了遮挡情况下重建新视角局部目标丢失现象，同时更好地保持新视角的边缘信息，获得更优的重建效果。

关键词：超分辨率重建；光场；卷积神经网络；自适应；全局视角；局部视角

中图分类号：TP391.41　　doi: 10.3969/j.issn.1001-3695.2017.12.0835

Light field super-resolution using global and local multi-views

Deng Wu, Zhang Xudong, Xiong Wei, Wang Yizhi

(*School of Computer & Information*, *Hefei University of Technology*, *Hefei* 230009, *China*)

Abstract: The spatial and angular resolution are comparatively low owing to the limitation of the resolution of its pixel sensor which is due to design reasons. To address the above problem, this paper presented a new super-resolution method based on global and local views to simultaneously enhance both the spatial and angular resolutions of a light field image. First of all, it adaptively chose local views according to the position of the novel view, and then used spatial super-resolution convolution neural network to get high resolution sub-images such as global-view and local-view images, subsequently it extracted and fused the depth and color features from the image obtained by backward warping the input views. Then put the results into the angular super-resolution network which produced a novel view image as the final result. The experimental results demonstrate that the reconstruction effect of this method is superior to the state-of-the-art's. And in peak signal to noise ratio get 3dB promotion and structural similarity image measure upgrade about 0.02. In addition, it is not only able to improve spatial and angular resolution of multi-view images, but also to maintain edge information of the novel view, thus effectively avoiding the loss of objects and yielding a better reconstruction effect.

Key words: super-resolution; light field; convolution neural network; adaptive; global views; local views

0 引言

随着光场成像技术的发展，光场相机[1]作为一种新型的多视角成像设备成为计算成像领域的研究热点。光场相机的基本原理是在主透镜和像素传感器中间插一个微透镜阵列，通过一次曝光可以获得 2D 空间信息和 2D 角度信息。相比于传统相机，光场相机具有先拍照后聚焦的独特优势，在显著性检测[2]、深度估计[3]、三维重建[4]等方面得到广泛应用。但是由于光场相机结构和像素传感器分辨率的限制，光场数据多出的 2D 角度信息是以牺牲一定的空间分辨率为代价的，两者之间存在一个折中。目前，光场商业相机如 Lytro、Lytro illum[5]和 Raytrix 都具有较低的角度分辨率，且其空间分辨率也低于传统相机,导致光场稀疏采样问题[6]，因此，同时提高光场图像的空间分辨率和角度分辨率对于充分发挥光场相机的优势，具有重要意义。

由于光场相机采集到的光场信息具有较多的冗余信息和先验信息，这使得利用超分辨率重建技术(super-resolution reconstruction，SR)来提高光场图像的分辨率存在可能。光场的超分辨率重建分为空间超分辨率重建和角度超分辨率重建。2009 年，Bishop 等人[7]利用深度图和精确的相机内部参数对光场相机逐层建模，并在贝叶斯框架下应用盲反卷积方法重建高

收稿日期：2017-12-18；修回日期：2018-02-07　　基金项目：国家自然科学基金资助项目（61403116）；中国博士后基金资助项目（2014M560507）

作者简介：邓武（1992-），男，安徽合肥人，硕士研究生，主要研究方向为计算机视觉、光场成像技术（dengwu3801@mail.hfut.edu.cn）；张旭东（1966-），男，安徽合肥人，教授，博士，主要研究方向为机器视觉、传感器技术、智能信息处理及其相关应用系统等；熊伟（1991-），男，湖北公安人，硕士研究生，主要研究方向为计算机视觉、光场成像技术；汪义志（1992-），男，湖北黄冈人，硕士研究生，主要研究方向为光场成像技术.

分辨率光场全聚焦图像。然而，该方法对微透镜图像直接进行重建，忽略了光场图像的角度分辨率。2012 年，Mitra 等人[8]分析了四维光场图像块的低秩性和图像块视差之间的关系，提出以混合高斯模型作为先验来提高光场图像的分辨率。但该方法仅适合漫反射场景，并且图像块需要取得足够小，从而忽略深度不连续的影响。2014 年，Wanner 等人[9]提出在全变分框架下，通过计算光场 EPI(epipolar plane image)获取全视角图像的深度信息，将贝叶斯框架下的 MAP 估计作为数据项，全变分模型作为能量函数的先验模型，对四维光场数据进行空间分辨率和角度分辨率的重建。其重建对象是所有的光场多视角图像，由于不需要精确的几何关系和相机内部参数，在一定程度上损失了重建图像的边缘信息。

近年来，深度学习尤其是卷积神经网络(convolutional neural network，CNN)成功应用于图像超分辨率重建[10~13]中。2014 年，Dong 等人[10]首次提出了基于卷积神经网络的图像超分辨率重建方法，分析了传统的稀疏编码和深度学习的 SR 技术的等价性，将 CNN 用于单幅图像的超分辨率重建。2016 年，Yoon 等人[11]首次提出的基于 CNN 对光场图像进行空间、角度超分辨率重建，利用了待重建视角周围的相邻视角，只能在相邻多视角图像之间合成新视角图像，不能在任意位置合成新视角。2016 年，Nima 等人[13]提出的基于 CNN 的光场图像角度分辨率重建方法，通过级联卷积神经网络分别估计新视角处的深度信息和颜色信息，来合成光场新视角图像。但该方法未提高新视角的空间分辨率，且仅使用四幅角点视角作为输入，对于复杂场景如遮挡，当角点视角对应的映射图像缺失有效信息时，会造成合成的新视角图像局部信息丢失。

针对上述问题，本文提出一种融合全局与局部视角的光场超分辨率重建算法，可以同时提高光场图像的空间分辨率和角度分辨率。鉴于商业光场相机采集的光场图像，其相邻多视角之间具有较大相关性，利用待重建新视角周围的视角信息，采用全局视角和局部视角相结合的方式，可以解决遮挡情况下局部信息丢失问题。实验结果在定性和定量两个方面都验证了，本文提出的超分辨率重建算法在结果图像中能够保持较好的效果。

1　光场相机成像原理

空间中，每一个点和每一个方向的辐射函数的总和就是光场[14]。微透镜阵列式光场相机成像原理如图 1 所示，在主透镜和像素传感器之间插入一个微透镜阵列，物体上某一点反射的光线经主透镜聚焦到微透镜阵列上，然后通过微透镜阵列将光线按照入射方向分散到像素传感器上，因此像素传感器在记录场景空间信息的同时也记录了角度信息。

与传统相机不同的是，微透镜阵列中每一个微透镜在像素传感器上形成对应的子孔径图像，每一个子孔径图像都记录了来自目标物某一点不同方向的光线，从所有子孔径图像中选取同一方向的像素，即可得到该方向的完整图像，从而可以获得多视角图像。如图 2 所示，每一幅视角图像表示从 st 平面上所有同方向点到 uv 平面上对应点的光线的集合。

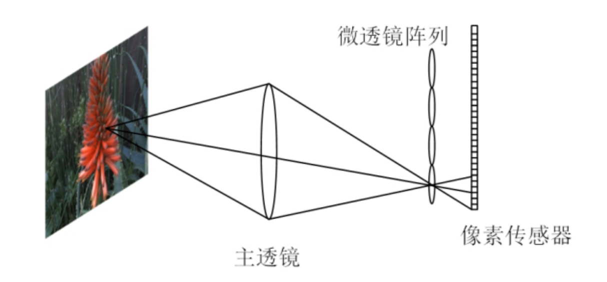


图 1　光场相机的成像模型

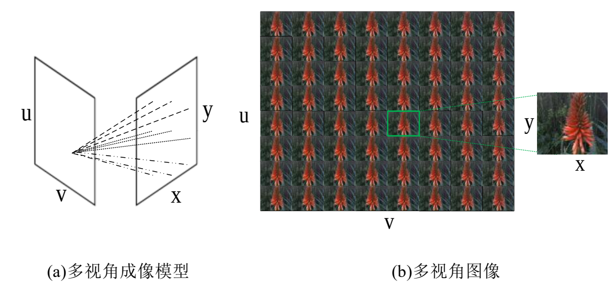


图 2　光场多视角图像

由于像素传感器的分辨率是固定的，每个微透镜记录光线对应相同位置不同角度的场景图像，这使得像素传感器的大部分像素用来记录方向信息，获得较大角度分辨率，导致空间分辨率较低。光场相机通过牺牲空间分辨率获取角度信息，其空间分辨率和普通相机相比低很多，角度分辨率也达不到实际应用需求。因此，本文通过超分辨率重建的方法，提高光场图像的空间分辨率和角度分辨率。

2　本文方法

本文的光场超分辨率重建旨在合成具有高空间分辨率的任意位置新视角图像。现有新视角合成方法都是通过已有视角来合成一个新的视角，主要分为：a)利用所有的光场多视角图像来重建一幅新视角图像，其重建新视角图像时间效率较低，平坦边缘区域容易产生阶梯效应；b)利用相邻局部多视角图像来重建一幅新视角图像，由于光场相机基线窄，局部视角方向信息很小，未考虑新视角位置的不同，局部视角选取问题，导致重建结果局部位置噪声较大，边缘模糊；c)利用全局视角来重建一幅新视角图像，时间效率较高，重建结果较好，但在遮挡情况下，会造成新视角图像局部信息丢失现象。综合考虑时间效率和遮挡情况，本文提出融合全局与局部视角的光场超分辨率重建算法。利用全局视角视差大，包含的角度信息大的特点，可以合成全局视角范围内任意位置的多视角图像；同时，结合局部视角和待重建新视角的相关性，可以解决遮挡情况下新视角图像局部信息丢失问题并且合成一幅新视角图像的效率较高，在选取局部视角时，根据局部视角和待重建新视角的相对位置关系，自适应选择局部视角。图 3 所示为使用不同视角的重建

结果，其中图 3(a)左上角为全局与局部视角示意图，每个方框代表一幅多视角图像，红色框表示全局视角，绿色框表示待重建新视角，蓝色框表示局部视角,右下角为局部放大真值；图 3(b)为使用相邻局部多视角的重建结果，图 3(c)为使用所有多视角的重建结果，图 3(d)为使用全局多视角的重建结果，图 3(e)为采用全局与局部视角的重建结果，可以看出，本文的算法的结果较好，避免了仅使用全局、局部和所有视角造成的局部信息缺失、边缘模糊和阶梯效应等问题。本文的算法整体框架如图 4 所示。为了合成高分辨率新视角图像，首先使用空间超分辨率卷积神经网络提高输入多视角图像的像素分辨率，得到高分辨率多视角图像；然后利用角度超分辨率卷积神经网络分别估计待重建新视角处的深度信息和颜色信息，通过融合全局视角和局部视角信息，优化其深度估计和颜色估计的精度，从而合成高质量的新视角图像。

本文算法主要步骤为：a)对光场原始数据进行预处理操作，解码重构出多视角图像，获取低分辨率全局视角图像和局部视角图像；b)为了得到高分辨率的全局视角图像和局部

视角图像，采用卷积神经网络提高全局视角图像和局部视角图像的空间分辨率；c)将步骤 b)得到的高分辨率全局视角图像和局部视角图像映射到待重建新视角位置，从映射图像中提取深度特征和颜色特征，利用角度超分辨率卷积神经网络，分别估计待重建新视角的深度信息和颜色信息，最终合成新视角图像。

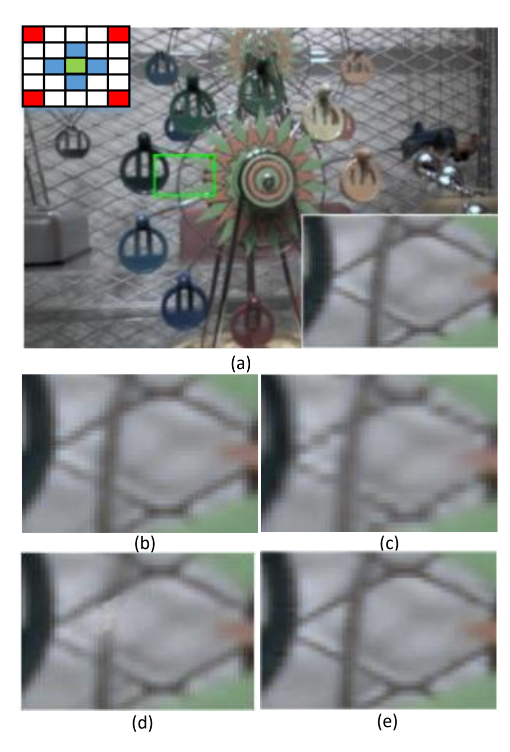


图 3　使用不同视角重建结果

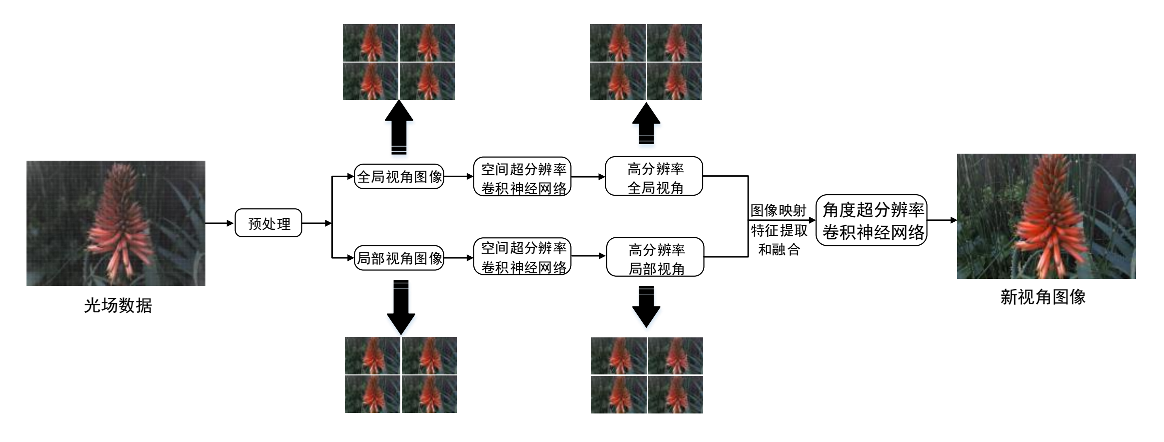


图 4　本文算法整体框架

2.1　光场数据预处理

采用 Lytro illum 相机获取的光场数据包含目标物同一位置不同方向的信息，为了重构出该目标物不同方向的多视角图像，通过 lytro-power-tools 解码原始光场数据，获得多视角图像。由于光场基线窄，光场相邻多视角图像之间的视差较小，其相关性较大，因此本文充分利用相邻光场多视角图像之间的相关性信息，融合全局视角和局部视角信息，重建出高分辨率的新视角图像。根据和待重建新视角的相对位置关系，分以下九种情况自适应选择局部视角。

当 $dx \le \frac{w}{4}$，$dy \le \frac{h}{4}$ 时，

$$L(i,j)=\{C(i-1,j),C(i+1,j),C(i,j\text{-}1),C(i,j+1)\}$$

(2)当 $\frac{w}{4}<dx\le\frac{w}{2}$，$\frac{h}{4}<dy\le\frac{h}{2}$ 且 $N_x<C_x(i,j)$，$N_y<C_y(i,j)$ 时，

$$L(i,j)=\{C(i-1,j-1),C(i-1,j),C(i,j),C(i,j-1)\}$$

(3)当 $\frac{w}{4}<dx\le\frac{w}{2}$，$\frac{h}{4}<dy\le\frac{h}{2}$ 且 $N_x<C_x(i,j)$，$N_y>C_y(i,j)$ 时，

$$L(i,j)=\{C(i-1,j),C(i-1,j+1),C(i,j+1),C(i,j)\}$$

(4) 当 $\frac{w}{4}<dx\le\frac{w}{2}$，$\frac{h}{4}<dy\le\frac{h}{2}$ 且 $N_x>C_x(i,j)$，$N_y<C_y(i,j)$ 时，

$$L(i,j)=\{C(i,j-1),C(i,j),C(i+1,j),C(i+1,j-1)\}$$

(5) 当 $\frac{w}{4}<dx\le\frac{w}{2}$，$\frac{h}{4}<dy\le\frac{h}{2}$ 且 $N_x>C_x(i,j)$，$N_y>C_y(i,j)$ 时，

$$L(i,j)=\{C(i,j),C(i,j+1),C(i+1,j+1),C(i+1,j)\}$$

(6) 当 $dx \le \frac{w}{4}, \frac{h}{4} < dy \le \frac{h}{2}$ 且 $N_y < C_y(i,j)$ 时，

$$L(i,j)=\begin{Bmatrix} C(i-1,j-1),C(i,j-1),C(i+1,j-1), \\ C(i-1,j),C(i,j),C(i+1,j) \end{Bmatrix}$$

(7) 当 $dx \le \frac{w}{4}, \frac{h}{4} < dy \le \frac{h}{2}$ 且 $N_y > C_y(i,j)$ 时，

$$L(i,j)=\begin{Bmatrix} C(i-1,j-1),C(i,j),C(i+1,j), \\ C(i-1,j+1),C(i,j+1),C(i+1,j+1) \end{Bmatrix}$$

(8) 当 $\frac{w}{4} < dx \le \frac{w}{2}, dy \le \frac{h}{4}$ 且 $N_x < C_x(i,j)$ 时，

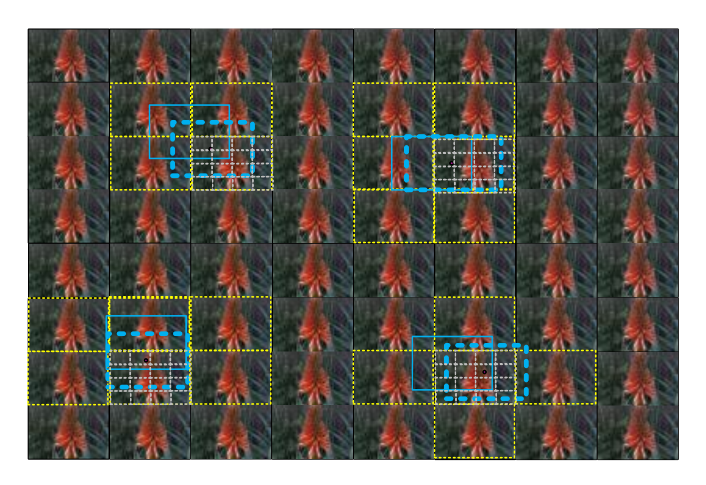
$$L(i,j)=\begin{Bmatrix} C(i-1,j-1),C(i-1,j),C(i-1,j+1), \\ C(i,j+1),C(i,j),C(i,j-1) \end{Bmatrix}$$

(9) 当 $\frac{w}{4} < dx \le \frac{w}{2}, dy \le \frac{h}{4}$ 且 $N_x > C_x(i,j)$

$$L(i,j)=\begin{Bmatrix} C(i,j-1),C(i,j),C(i,j+1), \\ C(i+1,j+1),C(i+1,j),C(i+1,j+1) \end{Bmatrix}$$

其中：N_x 表示待重建新视角中心点的横坐标，N_y 表示待重建新视角中心点的纵坐标，$C(i,j)$ 表示距离待重建新视角中心点坐标最近的多视角，$C_x(i,j)$ 表示距离待重建新视角中心点坐标最近的多视角中心点的横坐标，$C_y(i,j)$ 表示距离待重建新视角中心点坐标最近的多视角中心点的纵坐标，$dx=|N_x - C_x(i,j)|$，$dy=|N_y - C_y(i,j)|$，w 表示多视角图像的宽，h 表示多视角图像的高。

如图 5(a)所示，图中蓝色框表示待重建新视角，黄色框表示局部视角。选取边长为多视角图像宽高的一半，中心点与多视角图像中心点重合的矩形区域作为局部视角选择参考窗，如图 5(b)绿色区域所示。



(a)全局视角与局部视角的选择情况　　(b) 局部视角选择参考窗

图 5　全局视角与局部视角

2.2　空间超分辨率卷积神经网络

为了提高光场数据预处理后得到的低分辨率全局视角和局部视角图像的空间分辨率，使用基于学习的空间超分辨率卷积神经网络来提高其空间分辨率，得到高分辨率的全局视角和局部视角图像。网络模型类似于文献[12],但输入输出为 n 幅全局视角图像或者局部视角图像(n 为 4 或者 6)，通过深层网络对低分辨率图像块到高分辨率图像块之间映射关系建模，结构如图 6 所示。该网络将阶数较小的卷积核多次级联，充分利用低分辨率和高分辨率较大图像区域的上下文信息，获取它们之间的相关性信息，从而得到低分辨率图像到高分辨率图像之间的映射关系。将低分辨率的全局视角和局部视角图像输入空间超分辨率卷积神经网络，获得高分辨率的视角输出。空间超分辨率卷积神经网络有 20 个权值层，该网络第一层包含 64 个卷积核，尺寸为 3×3×n，即每个卷积核和输入的四幅光场多视角图像做卷积操作，生成 64 个通道（特征图）；第二层至第十九层的配置参数都一样,包含 64 个卷积核，尺寸为 3×3×64，即每个卷积核和上一层的 64 通道中的 3×3 区域做卷积操作；最后一层包含 n 个卷积核，尺寸为 3×3×64，用来图像重建，得到高分辨率多视角图像。

前十九个卷积层后各连接一个 Relu 激活层。

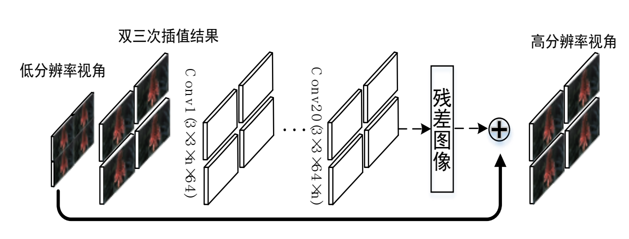


图 6　空间超分辨率卷积神经网络

2.3　角度超分辨率卷积神经网络

为了合成新视角，一般方法[9], [15-16]是先估算输入视角图像的深度信息，然后使用该深度信息将输入视角图像映射到待重建新视角位置，最后对映射图像进行操作（如加权）获得新视角图像。如图 7 所示，相机 c_i 从两个不同方向对物体面 Σ 拍摄成像于 Ω_i 平面的 x' 和 x 两点，利用深度估计算法得到输入多视角图像的深度图 $\boldsymbol{d}$，根据该深度图可知多视角图像平面 Ω_i 上的点 x 和物体表面 Σ 上的点 p 呈一一对应关系。由于光场相机超分辨率重建的新视角图像所在平面 Γ 与输入的多视角图像平面

Ω_i 往往在同一平面上，并且转移函数 $\boldsymbol{\tau}_i$ 跟深度图有关，其中 $\boldsymbol{\tau}_i(x)=x+\boldsymbol{d}_i(x)(c-c_i)$，因此将输入视角图像通过视角转移函数映射到待重建新视角位置，然后通过对映射图像操作获得新视角图像。

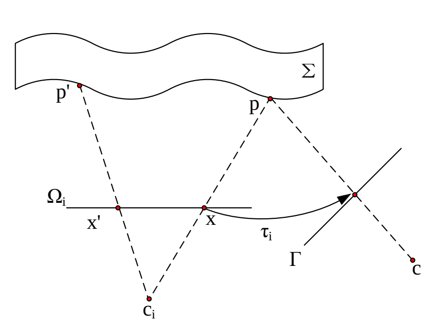


图 7　新视角合成示意图

Nima 等人将上述合成新视角方法分成深度估计和颜色估计两步，使用卷积神经网络分别对新视角处的深度信息和颜色信息建模，并用四个全局视角图像估计出待重建新视角位置的深度信息和颜色信息，从而重建出新视角图像。相比于现有合成新视角的方法而言，效果更好并且合成新视角图像的效率也有明显提高。然而对于复杂场景如遮挡，当全局视角对应的映射图像缺失有效信息时，会造成合成的新视角图像局部信息丢失。如图 8(a)所示，全局视角映射图像中丢失了柱子和叶尖信息，导致合成的新视角图像柱子和叶尖缺失。为了解决遮挡情况下出现的局部信息丢失问题，本文充分利用待重建新视角周围的多视角信息，采用全局视角和局部视角相结合的方式，融合全局视角和局部视角作为输入视角，使用卷积神经网络分别估计待重建新视角处的深度信息和颜色信息，最终合成高质量的新视角图像。如图 8(b)所示，局部视角映射图像中右上角和右下角图像包含了柱子以及叶尖信息，故最终合成新视角图像包含柱子和叶尖，解决了局部信息缺失问题。

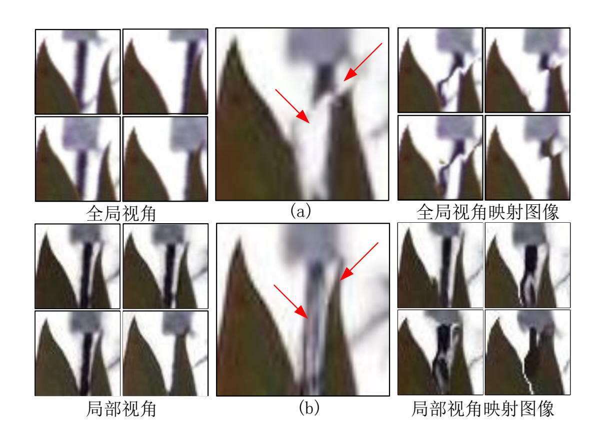


图 8　合成新视角结果对比

本文角度超分辨率重建的网络结构如图 9 所示，由输入视角图像 $Lp_{g_1},\ldots,Lp_{g_i},Lp_{l_1},\ldots,Lp_{l_n}$ 和新视角的位置 q，估计出新视角图像 Lq，即

$$Lq=f(Lp_{g_1},\ldots,Lp_{g_i},Lp_{l_1},\ldots,Lp_{l_n},q) \tag{1}$$

其中：Lp_{g_i} 表示全局视角图像，$i\in\{1,\ldots,4\}$，Lp_{l_n} 表示局部视角图像，$n\in\{1,\ldots,k\}$，其中 k 根据待重建新视角的位置选择 4 或 6，q 为新视角的角度坐标，f 定义为输入视角和新视角之间的映射关系。

2.4　深度估计模块

该模块的任务是估计待重建新视角图像的深度信息。为了估计深度信息，首先根据预设的深度等级 $d_1,\ldots,d_C$ 把全局视角图像 I_g 和局部视角图像 I_l 映射到新视角处：

$$\overline{L}_{p_t}^{d_c}(s)=L_{p_{g_i}}[s+(P_{g_i}-q)d_c]+L_{p_{l_n}}[s+(P_{l_n}-q)d_c] \tag{2}$$

其中：$t\in\{1,\ldots,N\}$，$c\in\{1,\ldots,C\}$，s 为 xy 方向的像素位置，q 为 uv 方向新视角图像的位置，p_{g_i} 和 p_{l_n} 分别为全局视角和局部视角在 uv 方向的位置，$\overline{L}_{p_t}^{d_c}(s)$ 为输入视角图像在每一个深度等级的映射图像。本文在新视角处[-21,21]像素范围内分成 100 个深度等级进行映射。利用 Tao 等人[17]的方法，从每一个深度等级的映射图像中提取散焦和一致性信息作为深度特征，即

$$M^{d_c}(s)=\frac{1}{N}\sum_{i=1}^{N}\overline{L}_{p_i}^{d_c}(s) \tag{3}$$

$$V^{d_c}(s)=\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\overline{L}_{p_i}^{d_c}(s)-M^{d_c}(s))^2} \tag{4}$$

通过连接每个深度等级的散焦信息和一致性信息 $\{M_g^{d_1},V_g^{d_1},\ldots,M_g^{d_C},V_g^{d_C},M_l^{d_1},V_l^{d_1},\ldots,M_l^{d_C},V_l^{d_C}\}$ 得到深度估计特征集 K。本文使用 100 个深度等级，每个深度等级包含全局与局部的散焦和一致性信息，因此深度特征集总共有 400 个特征向量。由于用全局视角和局部视角对应的新视角处映射图像提取深度特征来估计新视角的深度信息：

$$D_q=g_d(K) \tag{5}$$

其中：D_q 为待合成新视角图像的深度信息，g_d 定义为输入特征集和待合成新视角图像深度信息之间的映射关系。深度估计模块网络结构包含四个卷积层，其中第一层包含 200 个卷积核，尺寸为 7×7，第二层包含 100 个卷积核，尺寸为 5×5，第三层包含 50 个卷积核，尺寸为 3×3，第四层包含 1 个卷积核，尺寸为 1×1，前三层后连接 Relu 层。

2.5　颜色估计模块

该模块是对输入视角图像进行映射，利用映射图像、待重建新视角图像的深度信息及其位置坐标所组成的颜色特征集 $H=\{\overline{L}_{p_1},\ldots,\overline{L}_{p_N},D_q,q\}$，估计待重建新视角图像的颜色信息，其网络结构类似于深度估计模块。由于新视角图像的每一个像素，其深度信息和每一幅输入视角图像（包括全局视角和局部视角）的相应像素相对应，因此新视角图像的颜色信息可以由输入视角图像估计获得。根据深度估计模块估计得到的深度信息将输入视角图像映射到待重建新视角处，即

$$\overline{L}_{p_i}(s)=L_{p_{g_i}}[s+(p_{g_i}-q)D_q(s)]+L_{p_{l_n}}[s+(p_{l_n}-q)D_q(s)] \tag{6}$$

其中：$\overline{L}_{p_i}$ 为输入视角图像根据深度信息 D_q 得到的映射图像。

将颜色特征集输入颜色估计卷积神经网络，合成最终的新视角，即

$$L_q = g_c(H) \tag{7}$$

其中：g_c 定义为颜色特征集和估计新视角图像之间的映射关系。

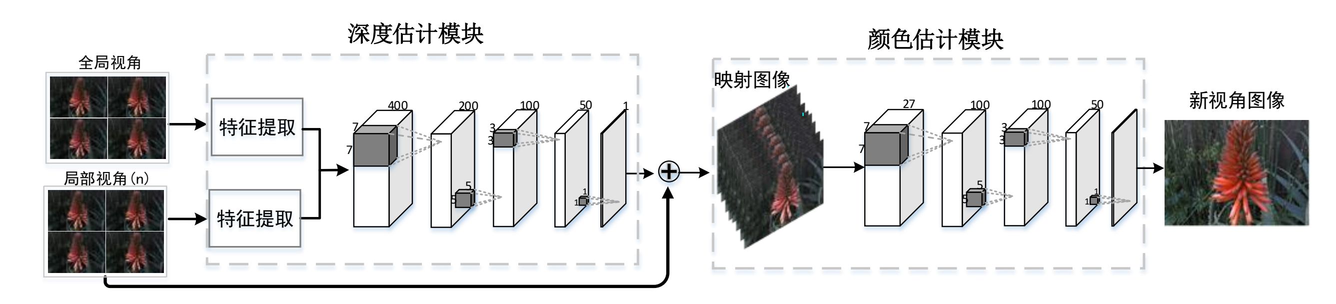


图 9 角度超分辨率卷积神经网络

3 实验结果及分析

3.1 实验设置

实验平台：实验使用的台式电脑配置如下：Intel-2620V4 CPU 2.1GHz x 8 cores，NVIDIA GeForce TITANX x 2 ，RAM 128G，Ubuntu 14.04 64 位操作系统，编译软件为 Matlab 2015b，Caffe，Matconvnet。

数据集：本文使用 standford 光场数据集[18]以及文献[13]的数据集作为训练集和测试集。这两个数据集都是由 Lytro illum 光场相机获取，其空间分辨率为 376×541，角度分辨率为 14×14，由于边缘子孔径图像受噪声、渐晕效应等影响，只采用中间 8×8 视角进行训练和测试。本文训练集包含 100 幅光场图片，包含自行车，汽车，大树，植物，人物等八类，测试集包含 30 张光场图片。对光场原始图像，按照尺寸为 60×60，步长为 16 像素提取图像块，因此本文的训练集数量有超过 10^6 个图像块。为了定量分析本文的超分辨率重建算法，实验中先删除光场图像的中心视角图像(5,5),并对其进行空间下采样，再以同样的采样因子（本文的采样因子为 2）上采样，然后将重建结果与原始的中心视角进行定量比较。

评价指标：为了验证本文方法的有效性，将本文方法与文献[9，11，13]的方法进行超分辨率重建的结果对比分析，并采用峰值信噪比（peak signal to noise ratio, PSNR）和结构相似性指数（structural similarity image measure, SSIM）进行定量评价。即：

$$\text{PNSR}=10\log_{10}\frac{W\times H}{\|\boldsymbol{x}-\boldsymbol{y}\|^2}$$

其中，W 、H 为图像的大小，$\boldsymbol{x}$ 为原始高分辨率图像，$\boldsymbol{y}$ 为重建后的高分辨率图像。

$$\text{SSIM}=\frac{(2\mu_x\mu_y+C_1)(\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_1)}$$

其中，μ_x 和 σ_x 分别为原始高分辨率图像的灰度平均值和方差，μ_y 和 σ_y 分别为重建后图像的灰度平均值和方差，σ_{xy} 为原始高分辨率图像和重建图像的协方差，C_1 、C_2 为常数。

3.2 网络训练

对于空间超分辨率卷积神经网络，对原始输入多视角图像先下采样再使用双三次线性插值的方法，以相同的缩放因子上采样得到低分辨率输入视角，原始多视角图像作为真值。对于低分辨率图像块和高分辨率真值图像块，本文在相同的区域随机裁剪尺寸大小为 41×41 的图像块作为输入和真值。在卷积操作之前，使用填充零的方法使得所有的特征图以及输出图像的尺寸大小保持不变，这样可以有效解决边缘像素的超分辨率重建问题。本文使用残差训练的方法训练该网络，对低分辨率和高分辨率图像之间的残差进行训练，并采用较高的学习率，加快收敛速度，同时采用可调节梯度裁剪策略[19]，调整梯度信息，有效避免训练过程中梯度消失和梯度爆炸的问题。

对于角度超分辨率卷积神经网络，由于现有真实场景光场数据集没有对应的深度真值，所以难以通过直接最小化估计的深度值和对应的深度真值来训练深度估计网络。本文通过最小化新视角图像和多视角真值图像之间的误差来训练角度超分辨率重建网络，能量损失函数为：

$$E=\sum_{k=1}^{3}(L_{q,k}-L_{q,k})^2$$

其中：$L_{q,k}$ 为合成的新视角图像，$L_{q,k}$ 为合成的新视角图像位置所对应的多视角真值图像，k 为 RGB 三通道。

本文使用 Caffe[20]训练网络，Matconvnet[21]来测试网络。空间超分辨率卷积神经网络使用均值为 0，标准差为 0.001 的高斯分布来随机初始化权重，动量参数为 0.9，权重衰减参数为 0.0001，学习率为 0.1。角度超分辨率卷积神经网络使用 Xavier 方法来随机初始化权重，求解器为 ADAM，其中 $\beta_1=0.9$，$\beta_2=0.999$，学习率为 0.0001。空间超分辨率卷积神经网络和角度超分辨率重建网络先分别单独训练，然后再将空间超分辨率卷积神经网络的输出作为角度超分辨率重建网络的输入，对整个网络进行端到端微调。

3.3 实验结果分析

为了验证本文算法的准确性，依次使用文献[9，11，13]方法以及本文方法对光场图像进行超分辨率重建。从提高角度分辨率，空间分辨率和重建效果以及算法时间效率上分析三种方法的超分辨率重建结果。

图 10 显示了使用本文算法进行角度超分辨率重建得到的新视角图像。图中第一行为输入多视角图像中的第一幅图像，

第二行为使用本文超分辨率重建算法获得的新视角图像。红色矩形框（图像右下角为其放大图像）表示输入图像与重建后图像相同位置同样大小的视图，可以看出重建后的图像与输入图像的视角不同。因此，超分辨率重建后获得一幅新视角图像，增加了光场图像的视角数，从而提高光场图像的角度分辨率。本文算法利用四个全局视角和局部视角，可以重建出全局视角范围内任意位置的新视角图像。



图 10 超分辨率重建后新视角图像

图 11 从定性的角度对比分析了文献[9]方法、文献[11]方法、文献[13]方法和本文方法在真实自然场景数据集上的超分辨率重建图像效果（为便于可视化，将重建图像部分放大显示）。

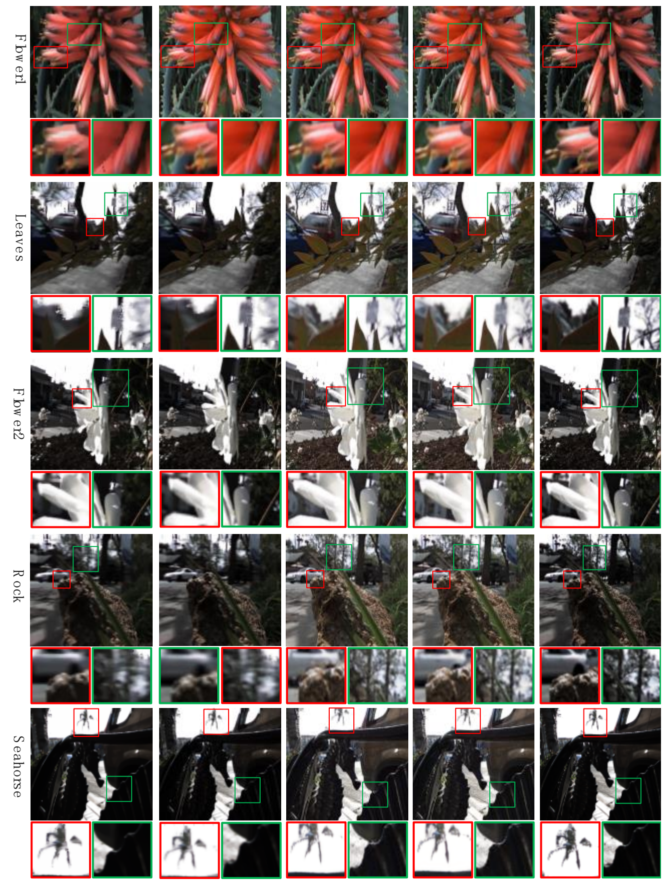


图 11 四种方法超分辨率重建结果定性结果

文献[9]方法使用了所有光场多视角图像来重建新视角，但不能很好的保持重建后图像边缘信息，导致重建后图像边缘模糊；文献[11]方法使用相邻多视角合成一幅新视角，对于复杂真实自然场景的合成效果较差，边缘噪声较大。文献[13]方法仅使用四幅全局视角图像来合成新视角图像，虽然很好的保持了边缘信息，但是对于一些遮挡场景，当其在待合成新视角处的映射图像缺失局部有效信息时，会造成合成的新视角图像局部目标丢失现象，比如 Leaves 场景中的柱子和 Rock 场景中的叶尖。本文方法采用了全局视角和局部视角相结合的方式，利用了局部视角的有效信息，可以弥补全局视角在待重建视角处映射图像局部信息丢失的缺陷,并且可以保持重建后图像的边缘信息。本文方法的重建效果优于文献[9]方法和文献[11]方法，不仅保持了重建后图像的边缘信息，而且有效的避免了局部目标丢失情况。

为了从定量的角度分析文献[9，11，[13]方法和本文方法的超分辨率重建结果，对真实自然场景数据集上进行超分辨率重建，分别采用峰值信噪比和结构相似性指数来量化，表 1 为定量比较结果，从表中可以看出，本文的方法要优于其他两种方法，与定性评价结果一致。

本文算法通过融合全局和局部视角信息，利用卷积神经网络提高光场图像的空间分辨率和角度分辨率，图 12 给出了不用方法在 standford 光场数据集上的峰值信噪比 PSNR 随迭代次数的变化趋势图。

从图中可以看出，传统方法不考虑迭代次数这个因素的影响，所以它们的 PSNR 保持不变，在图上呈直线表示，文献[11，13]最终的重建结果优于传统算法，而本文方法相对而言是最优的。

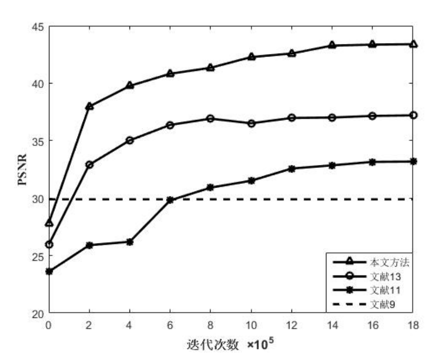


图 12 不同方法的重建结果对比曲线图

本文算法包括空间超分辨率重建网络和角度超分辨率重建网络两部分，根据网络复杂度的计算公式[22]：

$$O\left(\sum_{l=1}^{L} n_{l-1} f_l^2 n_l\right)$$

可以得到本文网络的复杂度为 5557204，文献[13]的网络复杂度为 2308404，文献[11]的网络复杂度为 251488。由于本文加入了局部视角，导致其输入的特征图(feature map)增大，从而网络复杂度增大，因此本文网络每次迭代的计算量较大，但由图 12 可知，本文网络在 14×10^5 接近收敛，总用时为 26.5 h，而文献[13]在迭代将近 12×10^5 收敛，总用时为 24.8 h，两个网络模型在大概相同的时间内收敛，但本文算法可以获得更好的

超分辨率重建效果。在时间效率上，表 2 给出了各算法重建一幅新视角图像的时间。

表 1 超分辨率重建结果的 PSNR 和 SSIM 比较

光场图像	文献[9]		文献[11]		文献[13]		本文方法	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Flower1	29.89	0.819	31.98	0.913	35.56	0.961	37.08	0.969
Leaves	22.68	0.810	26.32	0.892	27.81	0.933	32.65	0.967
Flower2	25.97	0.829	29.53	0.912	31.28	0.948	35.08	0.972
Rock	24.53	0.799	28.33	0.824	33.74	0.957	34.17	0.965
Seahorse	26.92	0.896	29.25	0.924	31.27	0.963	35.67	0.978

表 2 不同算法合成一幅新视角的时间效率

	使用所有多视角	使用局部多视角	使用全局视角	使用全局与局部相结合
效率	22.3s	16.6s	12.3s	12.5s

4 结束语

本文针对光场图像的空间分辨率和角度分辨率较低的问题，提出了融合全局与局部视角的光场超分辨率重建算法。通过融合包含信息最大的全局视角和根据和待重建新视角相对位置关系自适应选择的局部视角信息，可以合成全局视角范围内任意位置的新视角图像，并解决遮挡情况造成局部视角丢失的问题；同时为了提高新视角图像的空间分辨率，使用空间超分辨率卷积神经网络提高输入全局视角和局部视角图像的分辨率，从而得到高分辨率新视角图像。实验结果表明，本文方法可以快速合成一幅新视角，不仅保持更好的边缘效果，而且有效的避免局部目标丢失现象，获得相对更优的重建效果。

参考文献：

[1] Ng R, Levoy M, Brédif M, *et al.* Light field photography with a hand-held plenoptic camera [J]. Computer Science Technical Report, 2005, 2 (11): 1-11.

[2] Zhang J, Wang M, Lin L, *et al.* Saliency Detection on Light Field: A Multi-Cue Approach [J]. ACM Trans on Multimedia Computing, Communications, and Applications, 2017, 13 (3): 32.

[3] Tao M W, Srinivasan P P, Hadap S, *et al.* Shape estimation from shading, defocus, and correspondence using light-field angular coherence [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2017, 39 (3): 546-560.

[4] Vagharshakyan S, Bregovic R, Gotchev A. Light field reconstruction using Shearlet transform [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2018, 40 (1): 133-147.

[5] Georgiev T, Yu Z, Lumsdaine A, *et al.* Lytro camera technology: theory, algorithms, performance analysis [C]// Multimedia Content and Mobile Devices. International Society for Optics and Photonics. 2013, 8667: 86671J.

[6] 覃亚丽，张晓帅，余临倩. 基于低秩矩阵分解的光场稀疏采样及重构 [J]. 光学精密工程, 2017, 25 (5): 1171-1177.

[7] Bishop T E, Zanetti S, Favaro P. Light field superresolution [C]// Proc of IEEE International Conference on Computational Photography. 2009: 1-9.

[8] Mitra K, Veeraraghavan A. Light field denoising, light field superresolution and stereo camera based refocussing using a GMM light field patch prior [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2012: 22-28.

[9] Wanner S, Goldluecke B. Variational light field analysis for disparity estimation and super-resolution [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2014, 36 (3): 606-619.

[10] Dong C, Loy C C, He K, *et al.* Learning a deep convolutional network for image super-resolution [C]// Proc of European Conference on Computer Vision. Springer, 2014: 184-199.

[11] Yoon Y, Jeon H G, Yoo D, *et al.* Learning a deep convolutional network for light-field image super-resolution [C]// Proc of the IEEE International Conference on Computer Vision s. 2015: 24-32.

[12] Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1646-1654.

[13] Kalantari N K, Wang T C, Ramamoorthi R. Learning-based view synthesis for light field cameras [J]. ACM Tran on Graphics, 2016, 35 (6): 193.

[14] 尹晓艮，张晓芳，张伟超，等. 基于光场数字重聚焦的三维重建方法研究 [J]. 光电子·激光, 2015 (5): 986-991.

[15] Chaurasia G, Sorkine O, Drettakis G. Silhouette-Aware Warping or Image-Based Rendering [C]// Computer Graphics Forum, 2011, 30 (4): 1223-1232.

[16] Chaurasia G, Duchene S, Sorkine-Hornung O, *et al.* Depth synthesis and local warps for plausible image-based navigation [J]. ACM Trans on Graphics, 2013, 32 (3): 30.

[17] Tao M W, Hadap S, Malik J, *et al.* Depth from combining defocus and correspondence using light-field cameras [C]// Proc of IEEE International Conference on Computer Vision. 2013: 673-680.

[18] Raj A S, Lowney M, Shah R, *et al.* Stanford Lytro light field archive [J]. (2016) [2017-12-18]. http://lightfields. stanford. edu.

[19] Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks [C]// Proc of International Conference on Machine Learning. 2013: 1310-1318.

[20] Jia Y, Shelhamer E, Donahue J, *et al.* Caffe: Convolutional architecture for fast feature embedding [C]// Proc of the 22nd ACM International Conference on Multimedia. New York: ACM Press, 2014: 675-678.

[21] Vedaldi A, Lenc K. Matconvnet: Convolutional neural networks for matlab [C]// Proc of the 23rd ACM International Conference on Multimedia. New York: ACM Press, 2015: 689-692.

[22] He K, Sun J. Convolutional neural networks at constrained time cost [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. 2015: 5353-5360.